EX:No.1 DATE:25/01/25	**Implement Programs For Time Series Data Cleaning, Loading, And Handling Time Series Data And Pre-Processing Techniques**

AIM:

Write a program to implement time series data for import library, load data, Preprocessing and visualising.

OBJECTIVE:

- Load, clean, and analyze **US Water pollution data (2012-2021)**.
- Handle **missing values and outliers** for better accuracy.
- Identify **pollution trends** over time.
- Visualize pollution levels using **graphs and time-series plots**.
- Help researchers and policymakers make informed decisions.

BACKGROUND:

- Air pollution affects **health, climate, and environment**.
- Major pollutants: **PM2.5, CO, NO2, SO2, O3**.
- Poor Water quality leads to **respiratory diseases and global warming**.
- Analyzing historical data helps in **trend detection and policy-making**.

SCOPE OF THE PROGRAM:

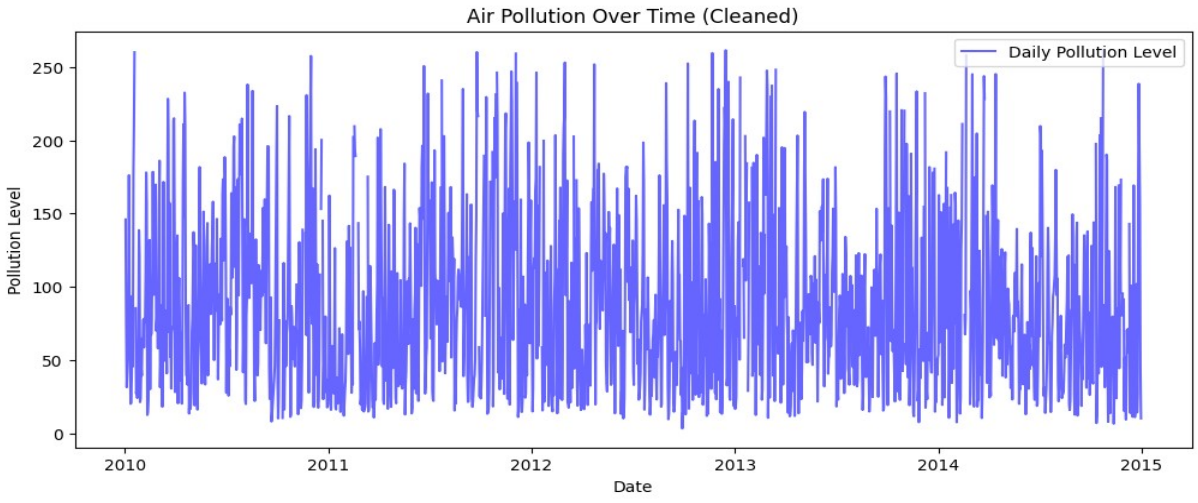
- **Data Processing & Cleaning**
- **Exploratory Data Analysis (EDA)**
- **Data Visualization**

CODE:

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

df = pd.read_csv("/content/Water_pollution.csv", parse_dates=["date"], index_col="date")

df = df[~df.index.duplicated(keep='first')]

df.fillna(method="ffill", inplace=True)

df.fillna(method="bfill", inplace=True)
```

```
df = df[['pollution_today']]

Q1 = df.quantile(0.25)

Q3 = df.quantile(0.75)

IQR = Q3 - Q1

df = df[~((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 * IQR))).any(axis=1)]

df = df.asfreq('D')

df_weekly = df.resample('W').mean()

plt.figure(figsize=(12,5))

plt.plot(df, label="Daily Pollution Level", color="blue", alpha=0.6)

plt.xlabel("Date")

plt.ylabel("Pollution Level")

plt.title("Air Pollution Over Time (Cleaned)")

plt.legend()

plt.show()
```

OUTPUT:



RESULT:

Thus, the program using the time series data implementation has been done successfully.